

FlyBase: improvements to the bibliography

Steven J. Marygold^{1,*}, Paul C. Leyland¹, Ruth L. Seal¹, Joshua L. Goodman², Jim Thurmond², Victor B. Strelets², Robert J. Wilson² and the FlyBase consortium[†]

¹Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK and ²Department of Biology, Indiana University, Bloomington, IN 47405, USA

Received September 13, 2012; Accepted October 3, 2012

ABSTRACT

An accurate, comprehensive, non-redundant and up-to-date bibliography is a crucial component of any Model Organism Database (MOD). Principally, the bibliography provides a set of references that are specific to the field served by the MOD. Moreover, it serves as a backbone to which all curated biological data can be attributed. Here, we describe the organization and main features of the bibliography in FlyBase (flybase.org), the MOD for *Drosophila melanogaster*. We present an overview of the current content of the bibliography, the pipeline for identifying and adding new references, the presentation of data within Reference Reports and effective methods for searching and retrieving bibliographic data. We highlight recent improvements in these areas and describe the advantages of using the FlyBase bibliography over alternative literature resources. Although this article is focused on bibliographic data, many of the features and tools described are applicable to browsing and querying other datasets in FlyBase.

CONTENT OF THE BIBLIOGRAPHY

FlyBase is a database of *Drosophila* genetic and genomic information, focusing on the model organism *Drosophila melanogaster*, but also including data on other *Drosophila* species and related drosophilids. The current content of the FlyBase bibliography reflects both the longevity of the database and the far longer history of *Drosophila* research: it comprises >200 000 references, including >87 000 research papers from >2400 different journals, with publication dates ranging from the 17th century through to the present day.

The FlyBase bibliography has always been defined fairly broadly in an effort to represent all research areas that use *Drosophila*. The basic criterion for inclusion of a research paper or other primary reference is that it reports on the biology of *Drosophila* as a direct experimental focus, whether whole organisms or the individual organs, cells or molecules contained therein. It follows that the bibliography contains research papers on a wide variety of topics, including classical genetics, modern genomics, molecular biology, evolution and ecology, bio-informatics and computational and theoretical biology. Articles that describe the use of *Drosophila* as a technology in support of some other research aim, such as using *Drosophila* cells to produce a non-*Drosophila* protein or using a *Drosophila* DNA sequence to isolate a gene from a non-*Drosophila* species, are excluded.

Other, non-primary publication types entered into the bibliography include reviews, conference reports, biographies and books (Table 1). The criteria for including these types of publication are less strict than those for research papers: essentially, *Drosophila* data should be reviewed or commented on, or information of general interest to the *Drosophila* research community should be reported. Reviews that are focused on other fields and make only one or two passing mentions of *Drosophila* data published elsewhere are not incorporated into the bibliography.

The contents of the FlyBase bibliography have been derived from a number of sources over the years (see flybase.org/static_pages/docs/data_sources.html for a full listing). Important sources have included several independent *Drosophila* reference collections (most notably the series of bibliographies compiled by Irwin Herskowitz between 1952 and 1983), the *Drosophila* Offprint Collection (originally collated by Michael Ashburner and maintained at the Department of Genetics, University of Cambridge), 'The Red Book' (1) and feeds from several bibliographic databases.

*To whom correspondence should be addressed. Tel: +44 1223 333170; Fax: +44 1223 333992; Email: sjm41@gen.cam.ac.uk

Present addresses:

Paul C. Leyland, Brnikat Ltd., Little Shelford, Cambridge CB22 5HJ, UK.

Ruth L. Seal, HUGO Gene Nomenclature Committee, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK.

Robert J. Wilson, Division of Developmental Biology, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

[†]The members of the FlyBase consortium are provided in the Acknowledgements.

Table 1. Publication types in FlyBase (as of FB2012_05 release)

Publication type	Number of references	Publication type	Number of references
Bibliography	102	Patent	396
Biography	284	Poem	7
Autobiography	10	Poster	80
Interview	34	Press release	13
Obituary	142	Research paper	87 447
Book	854	Retraction	16
Book review	190	Review	11 994
Comment	594	Sequence record	43 241
Compendium	9339	DNA/RNA sequence record	40 266
Edited book	2007	Protein sequence record	2975
Journal	2444	Species list	8
Conference abstract	37 638	Spoof	8
Conference report	300	Stock list	2972
Database	3	Supplementary material	487
Editorial	50	Teaching article	47
Erratum	344	Thesis	9248
Factual film	93	Unpublished	5701
Genome annotation	71	FlyBase analysis	121
Letter	356	Personal communication to FlyBase	5374
News article	74	Website	5
Newspaper article	61	White paper	6

Nowadays, practically all references in FlyBase are obtained through weekly queries of PubMed (see later). A separate list of FlyBase-authored references is also maintained and can be found at flybase.org/static_pages/docs/flybase-publications.html.

ORGANIZATION OF THE BIBLIOGRAPHY

Each reference in the bibliography is given a unique FlyBase reference (FBrf) identifier: an 'FBrf' prefix followed by seven digits, e.g. FBrf0123456. The number has no special meaning and simply reflects the order in which references are added to the bibliography.

The publication type of each reference is described using a controlled vocabulary that has been reviewed and revised in the past year (Table 1). Wherever possible, FlyBase publication types are now named and defined using the 'Publication Type' Medical Subject Heading (MeSH) terms used by PubMed (2). Most terms are self-explanatory, but a few are somewhat esoteric and worth expanding on here. 'FlyBase analysis' references describe changes to FlyBase data performed internally by FlyBase curators, such as edits to gene models. A 'personal communication to FlyBase', on the other hand, refers to observations submitted directly to FlyBase by a member of the research community that clarify published data, comment on other data in FlyBase or are stand-alone data files. Note that 'supplementary material' is classed as a distinct publication type in FlyBase. This is because supplementary data are not routinely curated by FlyBase, but are only captured (and labeled as such) when deemed particularly relevant. Also note that several publication types in Table 1 (such as conference abstracts, patents, sequence records and theses) are rarely used nowadays, as references of these types are no longer systematically added to the bibliography.

Most references in the bibliography have a 'parent publication', which for the majority of cases is a 'journal'. In other instances, the parent publication is an 'edited book', that is, a book containing several chapters on distinct topics, usually written by different authors and edited as a whole by one or more editors. Individual chapters within an edited book are then classified as independent references (usually 'reviews') in FlyBase with unique FBrf identifiers. This means that edited books are dissected into smaller portions that are easier for FlyBase curators and users to manage, which is especially useful where only a subset of chapters of a book concern *Drosophila*. In addition to a defined publication type ('journal', 'edited book' or the catch-all term 'compendium'), all parent publications in FlyBase are associated with a formal title and an abbreviated title (which mirror those used in PubMed where applicable), together with their International Standard Serial/Book Number (ISSN/ISBN). Indexing parent publications in this way ensures that they are all referred to uniquely and consistently within FlyBase, which helps prevent redundancy in the bibliography and is crucial for accurate querying of the data.

In addition to an FBrf ID, publication type and parent publication, several other standard citation fields are used to annotate references in FlyBase (Table 2). The PubMed identifier (PMID) and digital object identifier (DOI) are particularly important, as they are unique and global IDs that facilitate inter-database querying and cross-linking (see later). The 'Related publication(s)' field is also useful, as it links a research paper with its supplementary material or other related publication type, such as a specific commentary or erratum.

BIBLIOGRAPHY UPDATE PIPELINE

Previously, three different literature databases (BIOSIS, Zoological Records and PubMed) were searched regularly

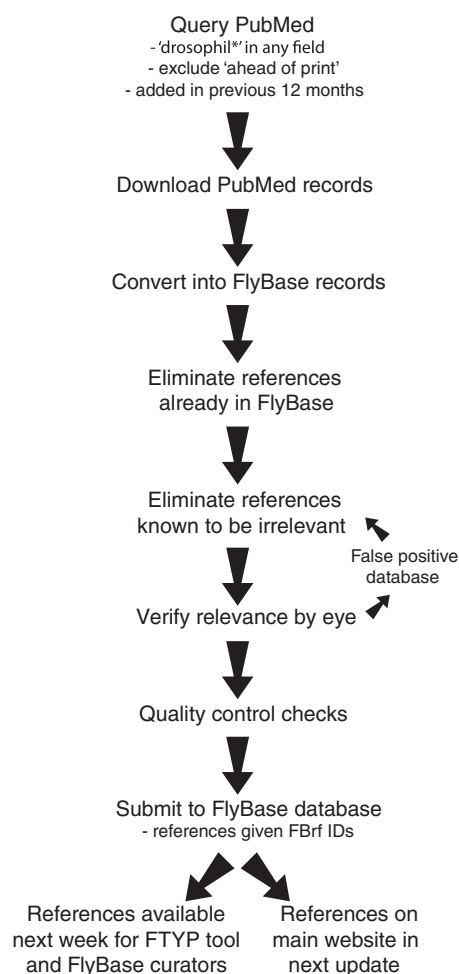
Table 2. Primary data fields used for a typical research paper in FlyBase

Field	Example
FlyBase reference ID	FBrf0212834
Publication type	Research paper
Author(s)	Weiss L.A., Dahanukar A., Kwon J.Y., Banerjee D. and Carlson J.R.
Year	2011
Title	The molecular and cellular basis of bitter taste in <i>Drosophila</i> .
Parent publication	Neuron
Volume	69
Issue	2
Page range/Article number	258–272
PubMed ID	21262465
DOI	10.1016/j.neuron.2011.01.001
Related publication(s)	Supplemental information (FBrf0213039)
Language of publication	English

to identify the maximal number of *Drosophila* references for the FlyBase bibliography. A drawback of this approach was that the same reference was often identified two to three times, resulting in the need for duplicate matching algorithms and, despite these, the introduction of redundancy into the bibliography. After PubMed began to index many more life science journals in 2006 (3,4), the additional references obtained via BIOSIS and Zoological Records were judged too few to merit the extra work involved in integrating three different database searches. Therefore, since 2008, references in FlyBase have been retrieved exclusively from PubMed. This more streamlined approach facilitated an increased frequency of bibliography updates, from just one to two per year before 2008 to the weekly updates of today.

The pipeline used to identify new *Drosophila* references and retrieve their citation data from PubMed has evolved during the past few years into an efficient and robust strategy for populating the FlyBase bibliography. The current semi-automated pipeline is summarized in Figure 1 (technical details are available on request). The search string ‘drosophil*’ identifies all references that mention *Drosophila* or drosophilids in their title or abstract, or have been associated with a *Drosophila*-related MeSH term by PubMed annotators. References marked in PubMed as ‘ahead of print’ are excluded from this search to avoid having to edit citation data in FlyBase (to reflect final volume and page numbers), and to prevent curation of any data that do not appear in the final version of a manuscript. Note that the weekly search is for references added to PubMed in the previous 12 months, rather than just in the past week. This ensures the identification of relevant references that have their publication status elevated from ‘ahead of print’ to ‘final format’, or that are annotated with a *Drosophila*-related MeSH term, a week or more after first appearing in PubMed.

The PubMed records identified by these criteria are then downloaded in XML format, and the pertinent citation data (Table 2) are extracted and used to create FlyBase records (the terms for publication type, journal abbreviation and publication language used in the PubMed records are converted to the matching terms used in FlyBase if necessary; other citation data are imported directly). Next, the PMIDs and DOIs of the current batch of

**Figure 1.** Bibliography update pipeline. See text for details.

references are automatically screened in order to remove (i) any relevant references that are already in the FlyBase bibliography, and (ii) any irrelevant, ‘false positive’ hits that were marked as such in previous updates. Common false positives include articles whose title/abstract states that the ortholog of a *Drosophila* gene is studied, or articles that have been annotated with a *Drosophila* MeSH term but that do not fulfill the criteria for inclusion in the

FlyBase bibliography (see earlier). The title/abstract and, if necessary, the full text of the remaining references are then rapidly checked by eye to remove any new irrelevant references, and to correct the publication types assigned by PubMed if necessary. (This manual verification step takes <30 minutes per week.) ~20% of the references identified in the original PubMed search are discarded during this step, highlighting the need for a manual check of relevance. The PMIDs of these irrelevant references are then added to a 'false positive' database to allow their automatic removal from subsequent updates. Finally, the validated records are checked computationally for basic formatting and syntax errors before being uploaded to the internal FlyBase database.

A few *Drosophila* references escape identification by the pipeline described earlier. This can happen when the title/abstract fails to mention '*Drosophila*' and *Drosophila*-related MeSH terms are not added within a year of publication; or the publication status of a reference fails to be updated from 'ahead of print' in PubMed within that same time frame; or if PubMed simply does not index the parent journal of a reference. Significant omissions are usually spotted swiftly, either by an interested FlyBase user or by a curator who is alerted to the missing reference when curating a related article, and are then added to the bibliography manually.

Approximately 55 new *Drosophila* references, of which ~45 are research papers, currently enter the FlyBase bibliography each week, usually within 1 week of being published in their final format. These references are then available for data curation. In the first instance, this involves sending an automated email to the corresponding authors of new research papers directing them to the FlyBase 'Fast-Track Your Paper' (FTYP) tool, which allows them to prioritize their article for further data extraction by FlyBase curators (5). Although the internal FlyBase database and the FTYP tool are updated with bibliographic details weekly, the FlyBase website is updated less frequently—approximately every 2 months at the present time. This means that new references, and any data associated with them in the meantime, appear on the FlyBase website 3–12 weeks after their integration into the internal database.

REFERENCE REPORT PAGES

Bibliographic data on the FlyBase website are presented in 'Reference Reports' (Figure 2, left panel). As with all Report pages in FlyBase, related data fields are grouped into sections, with the most important sections being permanently open, whereas others can be opened by clicking the '+' symbol on a section title or by clicking the 'Open All' button at the top of the page. The Reference Report page has recently been reorganized to present a cleaner and more focused view of the data—all key fields are now shown in a permanently open section at the top of the Report. Most fields within the Reference Report are self-explanatory and relate directly to the basic citation data shown in Table 2. However, certain derived features of the Reference Report deserve special mention.

Within the main 'Reference' section, the 'Citation' field is hyperlinked to the journal's webpage for that reference so that the full text of the article and any related content can be accessed directly. After this is the 'PubMed ID', which is hyperlinked to the respective PubMed report, and the full PubMed abstract of the reference. The DOI is also given, where available, and also links directly to the appropriate page at a journal's website. Any 'Related Publication(s)' are listed immediately underneath the 'Reference' section, organized by publication type, and are hyperlinked to their respective Reference Report page.

Reference Reports provide support for reference management software in two ways. First, an 'Export to RIS' link is provided at the end of each citation; clicking on this link downloads a tagged text file of citation data that can be imported into any standard reference manager. Second, if a user has Zotero (zotero.org) installed in their web browser, then a Zotero icon appears in the URL bar when viewing a Reference Report; clicking on this icon imports the citation directly into a Zotero library.

The 'Data from Reference' section at the bottom of the Reference Report is particularly useful and informative. Here, all genetic and molecular entities (genes, alleles, transcripts, etc.) that have been associated with the reference, either through the FTYP tool or by FlyBase curators, are shown and hyperlinked to their own respective Report page. This section thereby provides a useful overview of the content of any particular reference, as well as acting as a hub to explore associated data located on other Reports, such as gene ontology annotations, expression data or mutant phenotypes.

Field-by-field documentation for the Reference Report can be found by clicking on the 'Help' button at the top right of a Reference Report page, or by selecting 'Help' -> 'Report help' -> 'Reference Report' on the navigation bar of any FlyBase webpage.

REFERENCES IN OTHER REPORT PAGES

In addition to listing all associated entities on each Reference Report, reciprocal links are made in the 'References' section of each gene or allele Report page to all references that mention that entity (Figure 2, right panel). References are listed as full citations where there are relatively few of them associated with a gene or allele. However, where there are too many associated references to display them all comfortably, which is true for most characterized genes, a summary view is presented instead (Figure 2, right panel). In this view, links are given that generate a list of all references or of a specified publication type, together with separate sections that give the full citations of recent research papers and reviews published within the past 3 years. Wherever a full citation is listed, a link is provided that connects to the appropriate Reference Report page.

References are also found throughout FlyBase as short citation attributions (e.g. Weiss *et al.*, 2011) for all curated data statements. Again, these short citations are hyperlinked to their respective Reference Reports.

SEARCHING BIBLIOGRAPHIC DATA IN FLYBASE

The FlyBase bibliography can be readily searched to identify either a single reference of interest or to find a group of references that match a set of criteria. In either case, the easiest and fastest way to query bibliographic data is to use the QuickSearch tool on the FlyBase homepage (6). This tool was recently redesigned, and there is now a dedicated 'References' tab that allows the user to search all the key reference fields in any combination (Figure 3). The References tab has an intuitive interface similar to that used by other bibliographic databases and reference management software. Four search fields, 'Author(s)', 'Year(s)', 'Title/Abstract' and 'Journal', are

shown by default, though alternative search fields, including 'Publication type' and 'PMID/FBrf', can be selected (example search terms are shown as grey text in each field, which disappear when a real search string is entered). An autocomplete option is available for several of these fields, activated by clicking the box at the bottom of the panel, and standard Boolean operators can be used in the appropriate fields. Fuzzy matching is used so that terms containing diacritical marks, such as the author surnames Grönke, Léopold or Viña, are retrieved in searches irrespective of whether the search string or the database entry contains the diacritical mark. Additional documentation can be found by clicking on the 'QuickSearch help' link.

Figure 2 displays two example Reference Report pages from FlyBase (release FB2012_05). The left panel shows a 'Reference Report' for a specific entry, with the 'Genes' subsection open, listing various genes including *Gr5a* (highlighted with a red circle). The right panel shows a 'Gene Report' for *Gr5a*, listing references with counts in parentheses. A red arrow points from the *Gr5a* gene in the left panel to the *Gr5a* references in the right panel.

Figure 2. Example Reference Report pages (from FlyBase release FB2012_05). An example Reference Report is shown on the left; the 'Genes' subsection of the 'Data from Reference' section is open. The References section from an example Gene Report (*Gr5a*) is shown on the right (the other sections of the Gene Report are omitted for clarity); note that the number of references of each publication type is given in parentheses in each case. The reciprocal link between the Reference Report and the Gene Report is indicated by the red circle and arrow. (Blue text indicates hyperlinks; green-highlighted text indicates a new reference added in the current FlyBase release.)

QuickSearch

Simple Expression Phenotype GO References Data Type

°Author(s) : e.g. 'Smith NOT Johnson'

Year(s) : e.g. '2004-2008', '>2001', etc.

Title/Abstract : search Title and Abstract text

°Journal : e.g. 'Dev. Biol.'

°QuickSearch autocomplete: ☐ Note: Wild cards (*) can be added to your search term

QuickSearch help

Search

Figure 3. References tab of QuickSearch. See text for details.

The QueryBuilder tool (6) allows more powerful searching of reference-based data and is accessible from the homepage or via the Tools menu on the FlyBase navigation bar. In addition to the ability to search any reference field in any combination, this tool also permits querying across different data sets, such as searching for references associated with a particular gene, gene ontology term or phenotype. Figure 4 shows a more complex multi-leg example that finds non-review references that are associated with a gene of interest and were published in a specified range of dates. Template queries of reference data, which can be modified as necessary, are available through the QueryBuilder interface, together with full documentation and additional examples.

The TermLink tool (6) can be used to obtain a list of references annotated with a particular publication type or published in a particular language (i.e. the two bibliographic data fields that use a controlled vocabulary), or just to get an overview of the FlyBase bibliography from these perspectives. Although such lists are of limited use in themselves, they can be used as the starting point for further analysis or querying, as described in the next section.

REFERENCE HITLISTS, ANALYSES AND BULK DOWNLOADS

Unless a single matching reference is identified, the output of a QuickSearch, QueryBuilder or TermLink search is a 'hitlist' (Figure 4, left panel). Reference hitlists comprise seven data columns that correspond to the citation fields of authors, year of publication, title, journal, volume number, page range and publication type. The hitlist can

be easily reordered according to any of these fields by clicking on the arrows next to the column titles. Individual entries can be selected/deselected by clicking on the appropriate tick box in the first column. As with individual Reference Reports, Zotero users will see a Zotero icon appear in the URL bar when viewing a hitlist of references that allows direct import of all citations in the list.

Limited processing of a hitlist can be conducted by clicking on the 'Results Analysis/Refinement' button at the top of the page. For references, the frequencies of individual authors, years of publication, journals or publication types of the selected entries can be assessed (Figure 4, right panel). From this view, clicking a number in the 'Related records' column will display the references of that particular subcategory, effectively refining the original hitlist. (Searches and analyses similar to these can reveal global trends in *Drosophila* publishing, such as that the journal 'PLoS ONE' has published the most *Drosophila* research papers during the past 5 years, or that the annual number of *Drosophila* research papers has gradually risen from 1622 in 1992, the year in which the 'Red Book' (1) was published and FlyBase began, to 2307 in 2011.) Further hitlist processing and download options are available by clicking the 'HitList Conversion Tools' button at the top-right of a hitlist page. For example, selected entries can be exported to a new QueryBuilder session, downloaded as citations in RIS format, or downloaded in customized form via the Batch Download tool.

Having obtained a list of references of interest, the Batch Download tool (6) facilitates the bulk download of any associated citation field(s), such as a list of

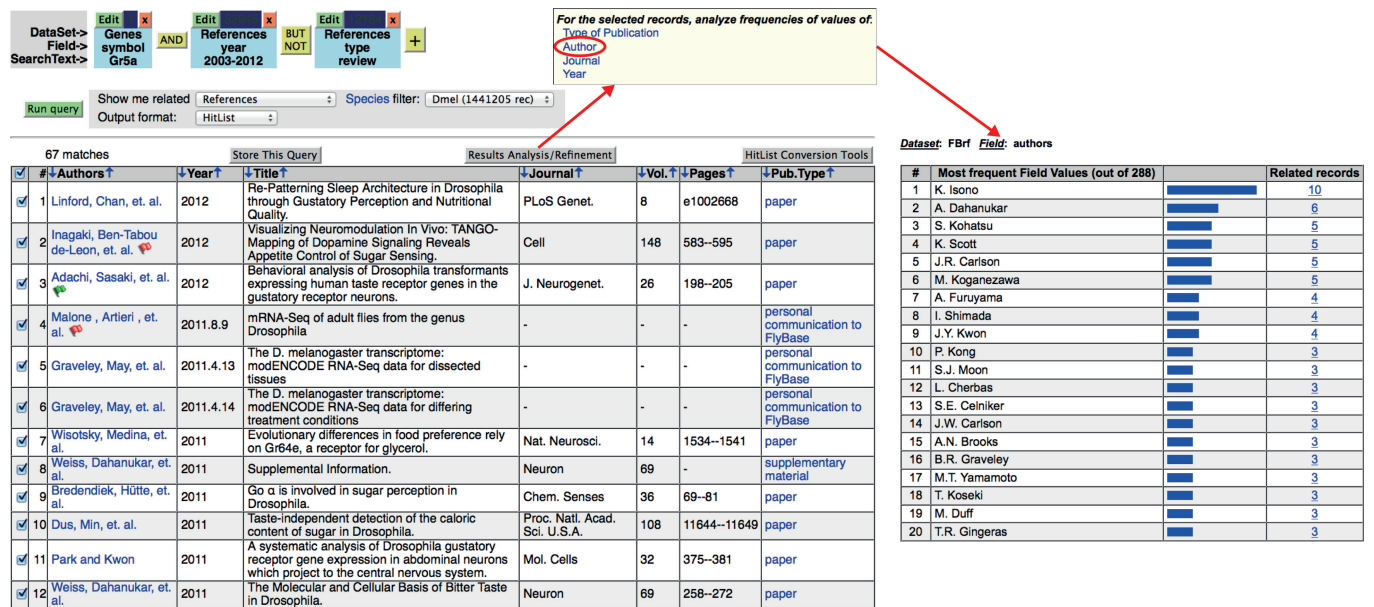


Figure 4. QueryBuilder, a reference hitlist, and frequency analysis (using FB2012_05). A QueryBuilder query for references, excluding reviews, published in the past 10 years that mention the gene *Gr5a* is shown. Below is the resulting hitlist of the first 12 matches, sorted in reverse chronological order. The 'Results Analysis/Refinement' pop-up menu is also shown, together with the frequency analysis of author names from the complete hitlist. (Blue text indicates hyperlinks; a green flag indicates a new reference added in the current FlyBase release; a red flag indicates that new data have been associated with a reference in the current release.)

PMIDs to allow further querying outside of FlyBase, or a file of abstract texts for reading offline. As described earlier, Batch Download can be reached via a hitlist, in which case the list of references, as FBrf IDs, is automatically transferred into the search box. Alternatively, Batch Download can be accessed directly from the homepage or via the Tools menu on the FlyBase navigation bar, in which case an FBrf/PMID list needs to be typed, pasted or uploaded. The citation data required for download are then chosen by selecting fields from an interface that mirrors a Reference Report page. Output options include an html table or a tsv file, viewed either directly in the browser or downloaded as a text file.

A pre-computed file of references that have an associated PMID, updated with each web release of FlyBase, has recently been made available via the 'Files' menu of the FlyBase navigation bar. It lists the FBrf, PMID, publication type, a short citation and the FlyBase release in which the PMID was added. This file is used to update the 'Textpresso for Fly' (7; textpresso.org/fly) search engine, but it is a generally useful list of all PMID-associated references in FlyBase that can be parsed to show references added in a specific FlyBase release.

CONCLUSIONS

The FlyBase bibliography is a well organized comprehensive and frequently updated literature resource that provides the infrastructure for data attributions in FlyBase. It also offers several advantages for reference-based searching over other public bibliographies, including *Drosophila* specificity, the inclusion of additional publication types such as personal communications and FlyBase analyses and the provision of manually curated links between each reference and the genes, alleles, transgenic constructs, etc. that feature within it. Several FlyBase tools can be used to interrogate bibliographic data, including a revamped reference-specific interface within QuickSearch. Hitlists of references can be sorted and further analyzed in a variety of ways, and standard or customized bulk downloads of bibliographic data can be easily obtained.

Questions about or suggested improvements to the bibliography, or any other area of FlyBase, are encouraged and can be submitted via the 'Contact FlyBase' link at the foot of any FlyBase webpage.

ACKNOWLEDGEMENTS

Michael Ashburner, Rachel Drysdale and Aubrey de Grey built the original FlyBase bibliography with assistance from Doreen Simpson (University of Cambridge Computing Service) and Jane Rosov (US National Library of Medicine). Michael Ashburner oversaw the updates for many years; subsequently, Eleanor Stanley and Gillian Millburn each performed the manual verification step for several years. The FlyBase Web Development

group coordinated the improvements to the display and searching of references on the website, Andy Schroeder wrote the code that generates the file of PMID-associated references, David Osumi-Sutherland helped with recent revisions to the publication type controlled vocabulary and Ray Stefancsik deputizes for running the weekly bibliography update. The current FlyBase consortium, members of which gave constructive comments on the manuscript, comprises: William Gelbart (PI), Kris Broll, Lynn Crosby, Gil dos Santos, David Emmert, Kathleen Falls, L. Sian Gramates, Beverley Matthews, Susan Russo, Andy Schroeder, Susan St Pierre, Pinglei Zhou and Mark Zytkevich (Harvard University, MA, USA); Nicholas H. Brown (PI), Boris Adryan, Helen Attrill, Marta Costa, Helen Field, Steven Marygold, Peter McQuilton, Gillian Millburn, Laura Ponting, David Osumi-Sutherland, Ray Stefancsik and Susan Tweedie (University of Cambridge, Cambridge, UK); Thomas Kaufman (PI), Kathy Matthews (PI), Josh Goodman, Gary Grumbling, Victor Strelets, Jim Thurmond and J.D. Wong (Indiana University, IN, USA) and Maggie Werner-Washburne (PI), Richard Cripps (PI) and Harriett Platero (University of New Mexico, NM, USA).

FUNDING

National Human Genome Research Institute at the US National Institutes of Health [P41 HG000739]; UK Medical Research Council [G1000968]; the Indiana Genomics Initiative; National Science Foundation through XSEDE resources provided by Indiana University. Funding for open access charge: NHGRI at the NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Lindsley, D.L. and Zimm, G.G. (1992) *The Genome of Drosophila melanogaster*. Academic Press, San Diego, CA.
2. Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S. et al. (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **40**, D13–D25.
3. Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **34**, D173–D180.
4. Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. et al. (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **35**, D5–D12.
5. Bunt, S.M., Grumbling, G.B., Field, H.I., Marygold, S.J., Brown, N.H., Millburn, G.H. and FlyBase Consortium (2012) Directly e-mailing authors of newly published papers encourages community curation. *Database*, bas024.
6. McQuilton, P., St Pierre, S.E., Thurmond, J.; FlyBase Consortium. (2012) FlyBase 101 - the basics of navigating FlyBase. *Nucleic Acids Res.*, **40**, D706–D714.
7. Müller, H.M., Kenny, E.E. and Sternberg, P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **11**, e309.